

Novice to Know-How Module Text

Course 7: Providing Access to Preserved Digital Content

Module 8: Creating Access Copies

The development of this course was funded by The National Archives (UK) as part of the "Plugged In, Powered Up" digital capacity building strategy.

1. Introduction.

When providing users with access to preserved digital content, we will not be granting direct access to our archived versions. There are many reasons for this, with keeping our collections secure from accidental or malicious damage being the most important. Instead, we will deliver to them copies specifically produced for access.

In this module we will examine the possible differences between archive and access copies, possible processes for creating access copies, considerations around file formats for access, and understanding what metadata and documentation might be needed by users. We will also provide examples of issues to consider for some common content types.

2. Archive Copies Versus Access Copies.

Creating access surrogates of preserved digital content, rather than providing direct access to the digital archive or repository, helps us to ensure that we do not open our digital collections up to security risks such as accidental deletion, changes to content, or malicious damage. It also provides us with the opportunity to consider if the original form of the digital content represents the most accessible version for our users.

Depending on the type of content in question, an access copy may be an exact replica of the original, or we may have new versions in different file formats according to the needs of our users. We may even decide to make several different access copies depending on the complexity of the digital content and our understanding of our differing users' needs. This is another area where information gathered during user needs analysis can help.

3. Choosing File Formats for Access.

When deciding what format(s) to use when providing access to preserved digital content, there are a number of issues to consider. It is important to think about each of the following:

User Skills.

We will hopefully have collected information during our user needs analysis about the skills for digital content use that our user groups might normally be expected to have. Some user groups will have less confidence in the use of technology than others, so we should try to provide access copies of digital content in formats they will be able to access without assistance where possible. This might mean creating an access copy in a widely used, easy-to-use format for one user group, while more advanced users could be offered a copy of the original, more complex format.

Availability of Software.

It is important to be aware of the availability of software to open, view, and use the format(s) we are considering using for access copies of preserved digital content. In particular if there are options that are freely available and easy to use. Particular file formats can only be opened with expensive proprietary software. During user needs analysis we can find out if our users will have access to the proprietary software. If they will not, we should try to identify a suitable substitute format that can be opened with a free viewer so as to not limit access.

Functionality.

When selecting formats for access copies, we also need to consider the different functionality offered by the various options. Some formats, and the software for using them, will offer more functionality than others, but will users need the full range of functionality? What is essential to allow users to be able to understand and use the digital content? User needs analysis should have provided some answers about how users wish to work with the digital content, and this should help us decide what functionality is essential, and what is not. If only certain functionality is required, we might be able to offer access copies in simpler, easier to access formats.

File Size.

We should also consider the size of the files produced by different formats and how this will impact access provision. If users will be downloading content, or copies will be transferred on removable media, huge files will slow down, or even prevent, access. We should, therefore, aim to offer access copies in formats that provide a balance between a good representation of the information content and original functionality, and the actual size of the file(s) produced from the preserved version.

4. When to Create Access Copies.

If we decide we need access copies in different formats, we will also need to consider at what stage in our digital preservation processes we will create the access copies. We might create them at the time of ingest, while processing the digital content for preservation. This may allow for more efficient workflows, and make it easier to provide downloadable versions for users, but will require us to store at least one extra copy, if not more, of all of our digital content.

An alternative approach is to create an access copy when the content is requested by the user. This will reduce storage requirements, but may increase the amount of time between a user making a request and receiving access to the content unless the process can be automated.

This decision of when to create access copies will be influenced by how much storage we have, how much digital content there is, if we can automate any processes, and how often we expect it to be accessed.

5. Thinking About Metadata and Documentation.

We should think about what metadata and documentation we might need to provide to help users access and use the content.

For most common types of digital content the basic information we might include in a catalogue entry will likely suffice: the file format(s), software used to create the content, and a brief description of the digital content. But more information might be required for more complex types of digital content.

A digital object might be composed of more than one file, so we may need to know about all of the required files to render the digital content onscreen, the relationships between the files and how to access them. For example, a Word document might include several links to content from other files, such as a special font or data from a spreadsheet that is included in a table.

6. Making Access Copy Decisions.

So far we have introduced issues that should be considered when creating access copies of digital content. Over the next few slides we will look at this in a little more detail using the following common types of digital content as examples:

- Digital Images,
- Text Documents,
- Audio-visual Content, and
- Computer Aided Design.

7. Example: Digital Images.

When deciding what access copies to create for digital images, one of the main concerns might be the size of the files. Many digital archives and repositories have their preservation copies of digital images saved in uncompressed TIFF format.

These files can easily be several hundred megabytes in size or larger, which could be prohibitively large for users on slower internet connections. For many users, a smaller JPG surrogate will provide a more accessible alternative whilst still meeting their reuse needs.

Some organizations will also use a charging model where smaller JPG versions are made available to users for free, but will charge a fee for access to a larger, higher quality, version in TIFF or other formats.

8. Example: Text Documents.

When deciding what formats to use for access copies of textual content there are a number of issues that may affect the decisions made.

While providing access to digital content in the original format is usually the preference, if we need to manage issues such as sensitive data this might not be possible. If you are providing limited access to a document with sensitive data you might create a read-only version in the original format or a locked PDF surrogate.

If you are providing onsite access only, then you will know what software users will have available and can select access formats accordingly. If, however, users can download content online, then you cannot guarantee what software they will have. For example, not all users will have a copy of the Microsoft Office Suite to open a .docx file. So, you may need to either offer the content in formats for which there are free software options, such as PDF or Open Document Format (ODF), or sign-post users to free software alternatives such as LibreOffice or Apache OpenOffice.

9. Example: Audio/Video.

When selecting access copy formats for audio-visual (AV) content there is usually a balance to be struck between file size, quality, and support for the format in commonly available software.

The most common formats for audio and video content are WAV and MPEG respectively. Although, when uncompressed these files can be prohibitively large when providing access. The FLAC (audio) and Matroska (MKV, video) formats can be used to provide high quality surrogates with much smaller file sizes, but we cannot always rely that users computers will support these formats. For example, both formats were only supported as standard by Microsoft operating systems from version 10 onwards. We might need to consider using lower quality alternatives, such as MP3 for audio.

User needs analysis may provide help with making these decisions. It allows us to understand what software our users have and for what purposes they are looking to use the content.

10. Example: Computer Aided Design.

Computer Aided Design (CAD) content is generally made using expensive, specialist software, that requires advanced skills to use. So, if we have non-specialist user groups, we cannot expect that they will have access to this software and/or the skills to use it.

CAD, therefore, is a good example of a content type where we might consider creating multiple types of access copies. For specialist users we can serve up a copy of the content in its original file format(s), and for non-specialists we can offer a simpler surrogate, perhaps as an image or PDF.

11. Module Wrap-Up.

In this module we have seen that we will not necessarily wish to provide access of copies of our preserved digital content in their preservation format(s). Instead, we will need to consider a range of issues that will influence what format(s) we will create access copies in. Those issues include:

- The skills our users have,
- The software available to them,
- What functionality is important for their reuse purpose, and
- The size of the files.

We will also need to consider what metadata and documentation to provide alongside access copies of content. And, finally, there are specific issues relating to different types of content that will influence our decisions about access copies.